JOURNAL OF APPROXIMATION THEORY **62**, 39–46 (1990)

Restricted Range Polynomial Interpolation

ALAN HORWITZ

Pennsylvania State University, Media, Pennsylvania 19063, U.S.A.

Communicated by Frank Deutsch

Received July 22, 1988; revised October 28, 1988

INTRODUCTION

Let f be a real-valued continuous function with $0 \leqslant f(x) \leqslant 1$ for all x in $[0, 1]$. We show that, for each positive integer n, there is a polynomial p of degree $\leqslant n$ that interpolates f at $n+1$ distinct points in $[0, 1]$, and such that $0 \leqslant p(x) \leqslant 1$ for all x in $[0, 1]$. (For certain f there is a restriction on the parity of n.) In other words, for *some* choice of $n+1$ distinct points in $[0, 1]$, the unique Lagrange interpolant to f at those points is bounded between 0 and 1. A similar result was proven by Briggs and Rubel [BR] for interpolation by non-negative polynomials, though our approach is different from their's. We use a perturbation of p^*, the best restricted range approximation (brra) to f in the uniform norm on $[0, 1]$, combined with the alternation theorem below. In fact it follows easily that if f is differentiable on $(0, 1)$, then p^* is a Hermite interpolant to f. Compare this to the unrestricted best approximation, which, by the Chebyshev Alternation Theorem, is a Lagrange interpolant to f.

In Section 2 we discuss restricted range interpolation where all the nodes are allowed to coalesce to form the nth partial sum of the Taylor series for f. We end the paper with some open questions.

1. MAIN RESULT

Let $H_n = \{$polynomials of degree $\leqslant n$ with $0 \leqslant p(x) \leqslant 1$ for all x in $[0, 1]\}$ and $\|f\| = \sup_{x \in [0,1]} |f(x)|$ for f in $C[0, 1]$. It is well-known that there is a unique polynomial p^* in H_n, called the brra to f, such that $\|f - p^*\| = \inf_{p \in H_n} \|f - p\|$. For fixed n and $f \notin H_n$ we define the following sets:

0021-9045/90 $3.00

$$E_+ = \{x \in [0, 1] : f(x) - p^*(x) = \|f - p^*\|\}$$
$$E_- = \{x \in [0, 1] : f(x) - p^*(x) = -\|f - p^*\|\}$$
$$C_+ = \{x \in [0, 1] : p^*(x) = 0\}$$
$$C_- = \{x \in [0, 1] : p^*(x) = 1\}$$

and $A = E_+ \cup E_- \cup C_+ \cup C_-$. Define the following function on A:

$$\sigma(x) = \begin{cases} +1 & \text{if } x \in E_+ \cup C_+ \\ -1 & \text{if } x \in E_- \cup C_-. \end{cases}$$

Important for our purposes are the following sets:

C = set of points in A where $p^* = f$. (Note that $C \subseteq C_+ \cup C_-$ if $f \not\equiv p^*$.) (1.1)

E = set of points in A where $p^* \neq f$, i.e., $E = A - C$. (1.2)

The key to our approach is the following theorem from approximation theory [T]. Here $l(x) \equiv 0$, $u(x) \equiv 1$, and our Chebyshev system is the algebraic polynomials.

ALTERNATION THEOREM (G.D. Taylor). *If $f \in C[0, 1]$ with $0 \leq f \leq 1$, then there exist $n+2$ points $x_1 < \cdots < x_{n+2}$ of A such that $\sigma(x_{i+1}) = -\sigma(x_i)$, $i = 1, ..., n+1$. We call $\{x_i\}_{i=1}^{n+2}$ an alternant for $f - p^*$.*

DEFINITION. Given a continuous function f on an interval I, if another continuous function g satisfies $f(t) = g(t)$ for some t in I, we say that g interpolates f at t. If $f - g$ also changes sign as x passes through t, we say that t is a crossover value for (f, g).

Our main result is

THEOREM 1. *Suppose $f \in C[0, 1]$ with $0 \leq f(x) \leq 1$ for all x in $[0, 1]$. Let n be a positive integer, where we assume*

(i) *n is even if $f(0) = f(1) = 0$ or 1*

(ii) *n is odd if $f(0) = 0$ and $f(1) = 1$, or $f(0) = 1$ and $f(1) = 0$.*

Then there exists a polynomial p in H_n which interpolates f at $n+1$ distinct points in $[0, 1]$.

Remarks. (1) If $0 < f(x) < 1$ for all x in $(0, 1)$, then it follows easily from the Alternation Theorem that the brra to f already interpolates f at

$n+1$ points in $[0, 1]$, i.e., no perturbation is necessary in this case. Theorem 1 is non-trivial when f takes on the values 0 and/or 1 in $(0, 1)$.

(2) If n is odd and f satisfies (i), or if n is even and f satisfies (ii), then we do not know if Theorem 1 still holds, except for $n \leqslant 3$ (see the remarks following the proof of Theorem 1).

(3) It is not hard to show that assumptions (i) and (ii) guarantee that 0 and 1 cannot both be in C if $f \not\equiv p^*$. For example, suppose $f(0) = f(1) = 0$ or 1. Now if $x_1 = 0$ and $x_{n+2} = 1$ are both in C, then by the definition of C we must have $p^*(0) = p^*(1) = 0$ or 1, so that 0 and 1 are either both in C_+ or both in C_-. This implies that $\sigma(0) = \sigma(1)$. But since n must be even by (i), by the Alternation Theorem $\sigma(1) = -\sigma(0)$, a contradiction. A similar argument shows that 0 and 1 cannot both be in C if $f(0) = 0$ and $f(1) = 1$, or $f(0) = 1$ and $f(1) = 0$. Finally, if either $0 < f(0) < 1$ or $0 < f(1) < 1$, then it follows easily that 0 or 1 must be in E.

Before proving Theorem 1 we need the following lemmas. For Lemmas 1 and 2, assume that f and g are continuous functions in some neighborhood N of $x = c$, and that f and g are never equal to d in $N - \{c\}$.

LEMMA 1. *Suppose $f(c) = g(c) = d$ and that (c, d) is either a local maximum or minimum point for both f and g. Then for r sufficiently close to* 1, *there is a crossover value t in $(c, c/r)$ for (f, g_r), where $g_r(x) = g(rx)$.*

Proof. Suppose that (c, d) is a local maximum point for both f and g. Since $g_r(c/r) = g(c) = d$, then for r close to 1, $g_r(c/r) > f(c/r)$ and $g_r(c) = g(rc) < d = f(c)$. The lemma then follows from the intermediate value theorem. The case when (c, d) is a local minimum point follows in the same way.

The argument we use to prove the following lemma is similar to the proof of case A of the theorem in [BR].

LEMMA 2. *Suppose that (c, d) is a local maximum (minimum) point for both f and g, and that $f - g$ is non-negative (non-positive) in N. Let t' and t'' be any two numbers in N with $t' < c < t''$, $f(t') \neq g(t')$, and $f(t'') \neq g(t'')$. Then for r sufficiently close to* 1, *there are crossover values for (f, g_r) in $(t', c/r)$ and $(c/r, t'')$.*

Proof. Without loss of generality, consider the case when (c, d) is a local maximum point for both f and g, and $f - g \geqslant 0$ in N. Choose $\varepsilon > 0$ so that $g(t') + \varepsilon < f(t')$ and $g(t'') + \varepsilon < f(t'')$, and choose r close to 1 so that $|g(x) - g_r(x)| < \varepsilon$ for $x = t', t'', f(c/r) < d$, and $t' < c/r < t''$. Then $g_r(t') < f(t')$, $g_r(t'') < f(t'')$, and $g_r(c/r) = g(c) = d > f(c/r)$. The lemma now follows again by the intermediate value theorem.

LEMMA 3. *Let $f \in C[0, 1]$, $0 \leqslant f \leqslant 1$, and let p^* be the best approximation to f from H_n, n a given positive integer. Assume that f takes on the values* 0 *and* 1 *only finitely often in* $[0, 1]$, *and that $f \not\equiv p^*$. Let $\{x_i\}_{i=1}^{n+2}$ be an alternant for $f - p^*$, and suppose there are two alternation points $x_i < x_k$ in E, with all of the alternation points between x_i and x_k in C. Then for r sufficiently close to* 1, *there are at least $k-i$ crossover values for $(f, p^*(rx))$ in (x_i, x_k).*

Proof. The lemma follows immediately if $k = i+1$, so we can assume $k \geqslant i+2$. Also assume, without loss of generality, that $p^*(x_i) < f(x_i)$. Let $S = \{x_{i+1}, ..., x_{k-1}\}$ and note that

$$p^*(x_{i+1}) = f(x_{i+1}), ..., p^*(x_{k-1}) = f(x_{k-1}). \tag{1.3}$$

Every point of S is either a local strict maximum or minimum for both f and p^*. (1.4)

Statement (1.4) follows for f since f cannot equal 0 to 1 infinitely often. Statement (1.4) follows for p^* since p^* cannot be a constant if $S \neq \varnothing$ (if p^* is a constant, then it follows easily that it must be 0. But then $f \equiv 0$ by the Alternation Theorem).

By (1.4) there are deleted neighborhoods of each of the points in S in which f and p^* cannot equal 0 or 1. We can then apply Lemmas 1 and 2 when needed.

Now by Lemma 1 (with $g = p^*$), for each point x_j of S, there is a crossover value for $(f, p^*(rx))$ in $(x_j, x_j/r)$ for r sufficiently close to 1. But this gives only $k-i-1$ crossover values, which is not enough for the lemma. To get the extra crossover value, we argue as follows.

We consider two cases.

Case 1. All of the points of S are crossover values for (f, p^*).

Suppose $p^*(x_k) < f(x_k)$. Then $k-i$ must be even by the Alternation Theorem since x_i and x_k are both in $E_+ \cup C_+$. Since all of the points of S are crossover values, $f - p^*$ must have at least $k-i-1$ sign changes in (x_i, x_k). But since $f - p^*$ is positive at both x_i and x_k, there must be precisely an even number of sign changes, and hence $f - p^*$ has at least $k-i$ sign changes in (x_i, x_k). This yields the extra crossover value, and the lemma now follows since crossover values are preserved under small perturbations. The case when $p^*(x_k) > f(x_k)$ follows in a similar fashion. This completes the proof of Lemma 3 when all the points of S are crossover values.

Case 2. Assume that at least one of the points of S is not a crossover value for (f, p^*).

Let m be the smallest positive integer, $m > i$, such that $x_m \in S$ and x_m is not a crossover value. Assume that $m - i$ is even, and thus $p^*(x_m) =$

$f(x_m)=0$, the case $m-i$ odd following in a similar fashion. We now have two possibilities.

(i) $p^*(x)<f(x)$ on $(x_m-\delta, x_m+\delta)-\{x_m\}$ for some $\delta>0$.

Since $x_{i+1}, ..., x_{m-1}$ are all crossover values for (f, p^*), $f-p^*$ has at least $m-i-1$ sign changes in $(x_i, x_m-\delta/2)$. But since $p^*(x)<f(x)$ on $(x_m-\delta, x_m)$ and $p^*(x_i)<f(x_i)$, the precise number of sign changes of $f-p^*$ in $(x_i, x_m-\delta/2)$ must be even, and hence must be at least $m-i$. Thus for r close to 1, $(f, p^*(rx))$ will have at least $m-i$ crossover values in $(x_i, x_m-\delta/2)$. As noted above we get crossover values to the right of x_m through x_{k-1} by Lemma 1. This gives a total of $k-i$ crossover values.

(ii) $p^*(x)>f(x)$ on $(x_m-\delta, x_m+\delta)-\{x_m\}$ for some $\delta>0$.

Then we just apply Lemma 2 with $g=p^*$ and $x_m-\delta<t'<x_m<t''<x_m+\delta$, to get that $(f, p^*(rx))$ has crossover values in $(t', x_m/r)$ and $(x_m/r, t'')$ for r close to 1. Again by Lemma 1 we obtain crossover values to the right of x_j for $j=i+1, ..., m-1, m+1, ..., k-1$. Since we just obtained two crossover values near x_m, this gives a total of $(m-i-1)+(k-m-1)+2=k-i$.

Note that since $(x_m, 0)$ is a local minimum point for both f and p^*, we can only use Lemma 2 for case (ii). Note also that we can always choose r, t', t'' appropriately so as not to count the same crossover value twice. This proves Lemma 3.

Proof of Theorem 1. Assume that $f-p^*$ has only finitely many zeros in $[0, 1]$ and that f takes on the values 0 and 1 at most a finite number of times (otherwise Theorem 1 follows immediately by taking $p=p^*$, or $p=0$ or 1, respectively). Let $\{x_i\}_{i=1}^{n+2}$ be an alternant for $f-p^*$ such that 0 and 1 are not both in C (such an alternant exists by the Alternation Theorem and Remark 3 following the statement of Theorem 1). Of course, the alternant may not contain 0 and/or 1 at all. Also, if all, or all but one, of the alternation points is in C, then we have at least $n+1$ interpolation points and the theorem is proved by taking $p=p^*$. So assume that there are at least two points in E, and let x_{p+1} be the first point in E and x_{q-1} the last such point. Let $B=\{x_1, ..., x_p\}$ and $D=\{x_q, ..., x_{n+2}\}$ (it could happen that B and/or D are empty). There are three cases to consider for the last alternation point.

Case 1: $x_{n+2}<1$.

Then by Lemma 1 we get a crossover value for $(f, p^*(rx))$ just to the right of each point in $B\cup D$ (if $x_1=0$ then x_1 is an interpolation point for all $r<1$). We also apply Lemma 3 to intervals of the form (x_i, x_k) where x_i and x_k are in E and any alternation points in between are in C. The total count of interpolation points must then be at least $n+1$ and Theorem 1 is proven in this case by taking $p(x) = p^*(rx)$, r close to 1. Note that the crossover value to the right of x_{n+2}, when x_{n+2} is in C, makes up for the possibly missing crossover value between x_{q-1} and x_q.

Case 2: $x_{n+2} = 1$ and $1 \in E$.

Then we argue as in Case 1, except that here the set D is empty and we do not need a crossover value to the right of x_{n+2}.

Case 3: $x_{n+2} = 1$ and $1 \in C$.

First, if $x_1 > 0$, then just consider $g(x) = f(1-x)$, use Case 1 for g, and map back. Second, if $x_1 = 0$ and $0 \in E$, then just apply Case 2 to g. This exhausts all possibilities by assumptions (i) and (ii) in Theorem 1, which guarantee that 0 and 1 cannot both be in C.

Remarks. (1) It is easily seen from the proof of Theorem 1 that assumptions (i) and (ii) can be replaced by the weaker assumption that 0 and 1 are not both in the set C. We preferred, however, to state Theorem 1 without any reference to the Alternation Theorem.

(2) For $n \leqslant 3$ Theorem 1 holds without assumptions (i) and (ii). For $n = 1$ this is trivial. Now suppose $n = 2$ and $p^*(0) = f(0) = 0$, $p^*(1) = f(1) = 1$, so that 0 and 1 are both in C. Then p^* is increasing on $[0, 1]$, and the two points in $A \cap (0, 1)$ must be in E (in fact, in $E_+ \cup E_-$). This implies that p^* itself interpolates f at three points in $[0, 1]$, and Theorem 1 is proved. If $n = 3$ and if $p^*(0) = p^*(1) = f(0) = f(1) = 0$ (so again 0 and 1 are both in C), then it is easy to show that there cannot be an alternation point in C between two points in E. But then p^* interpolates f at at least two points in $(0, 1)$, and Theorem 1 follows. This line of reasoning breaks down for $n \geqslant 4$.

2. Taylor Series

In this section we discuss restricted range interpolation when the $n+1$ interpolation nodes coalesce into one point c, and we then have the Taylor polynomial $s_n(x; c)$ of order n at c. In [BR] it is noted that if $f \in C^\infty[0, 1]$ is non-negative on $[0, 1]$, and if n is even, then it is possible to choose $c \in [0, 1]$ such that $s_n(x; c)$ is also non-negative on $[0, 1]$. We now show that this fails in the restricted range case for any $n \geqslant 2$. In fact we construct one f that works for all $n \geqslant 3$. So suppose f satisfies, for each $n \geqslant 3$,

$$f^{(n+1)} < 0 \qquad \text{on } [0, 1] \tag{2.1}$$

$$0 \leqslant f \leqslant 1 \qquad \text{on } [0, 1] \tag{2.2}$$

$$f(0) = 0, \qquad f(1) = 1, \qquad \text{and} \qquad f(x_0) = 1 \quad \text{for some } x_0 \text{ in } (0, 1). \tag{2.3}$$

It is not hard to construct such an f. For example, one can choose $f(x)=p(x)-ke^{rx}$ where p is a cubic polynomial and k and r are positive constants. Then f will automatically satisfy (2.1). We must choose p, k, and r to force f to satisfy (2.2) and (2.3). First let $g(x)=q(x)-e^x$, q a cubic, so that g has two local maximums and a local minimum in between. This can be done so that $f(x)=g(L(x))+c$ satisfies (2.2) and (2.3), for some linear function L and some constant c.

Now let $p(x)=s_n(x;c)$ for fixed $n\geqslant 3$ and c in $[0,1]$, and let $E(x)=f(x)-p(x)=[f^{(n+1)}(\zeta)/(n+1)!](x-c)^{n+1}$, ζ between x and c. Since $E(1)<0$ unless $c=1$, we must choose $c=1$—for otherwise $p(1)>1$. But if n is even, then $E(0)>0$, which implies $p(0)<0$. If n is odd, then $E(x_0)<0$, which implies that $p(x_0)>1$. Thus $0\leqslant f\leqslant 1$ on $[0,1]$, but no Taylor polynomial to f of degree at least three, expanded about c in $[0,1]$, has the same property. For $n=2$ we can choose f to satisfy (2.1)–(2.3), except we do not assume $f(x_0)=1$ for some x_0 in $(0,1)$, which is really only used when n is odd.

It is natural to ask what happens when $n=1$, i.e., must some tangent line to f be bounded between 0 and 1 on $[0,1]$, where we assume f is differentiable on $[0,1]$ with $0\leqslant f\leqslant 1$? We can prove the following partial result.

THEOREM 2. *Suppose* $0\leqslant f(x)\leqslant 1$, $f(0)=0$, $f(1)=1$, *and* $f'''(x)\neq 0$ *for all* x *in* $[0,1]$. *Then for some* c *in* $[0,1]$, *the tangent line at* $(c,f(c))$, $T_c(x)$, *satisfies* $0\leqslant T_c(x)\leqslant 1$ *for all* x *in* $[0,1]$.

Proof. We can assume f' is never 0, for otherwise there is a horizontal tangent line which does the job. Since $f(0)<f(1)$, we then have $f'>0$ on $[0,1]$. Now if f'' is never 0 in $(0,1)$, then we can choose the tangent line at $(0,0)$ if f is convex, and the tangent line at $(1,1)$ if f is concave. So suppose $(x_0,f(x_0))$ is an inflection point, where $0<x_0<1$ (Since f'' is monotone there is precisely one inflection point if f'' vanishes somewhere in $(0,1)$).

Case 1: $f''(x)<0$ for $x<x_0$.

Then we can choose $c=x_0$ for the following reason. Let $E(x)=f(x)-T_{x_0}(x)=(f''(\zeta)/2)(x-x_0)^2$, ζ between x and x_0. Then it follows immediately that $E(0)\leqslant 0$ and $E(1)\geqslant 0$, and since T_{x_0} is increasing, Theorem 2 is proven in this case.

Case 2: $f''(x)>0$ for $x<x_0$.

Note that we cannot have both $f'(0)>1$ and $f'(1)>1$, since then $f'(x)>1$ for all x in $[0,1]$, which contradicts, by the Mean Value Theorem, the fact that $(f(1)-f(0))/(1-0)\leqslant 1$. Now $T_0(x)=f'(0)x$, which implies that $T_0(1)=f'(0)$, and $T_1(x)=1+f'(1)(x-1)$, which implies that $T_1(0) = 1 - f'(1)$. Thus T_0 and/or T_1 must satisfy the conclusion of Theorem 2.

3. Open Questions

It is natural to ask how many of the interpolation points in Theorem 1 can be *specified in advance*. For non-negative interpolation this question was discussed in [H], where it was shown that roughly half the points can be fixed in advance if f is positive at those points. If $f(c) = 0$, $c \in [0, 1]$, then it is possible that there is no non-negative Lagrange interpolant to f when one of the interpolating points includes c. A similar negative result follows immediately for restricted range interpolation. For example, there is no non-negative quadratic (and hence no quadratic bounded between 0 and 1) that interpolates x^3 at three distinct points in $[0, 1]$, 0 included.

QUESTION 1. *How many interpolation points,* $(c, f(c))$, *in Theorem* 1 *can be specified in advance if* $0 < f(c) < 1$?

QUESTION 2. *Can assumptions* (i) *and* (ii) *be removed in Theorem* 1?

QUESTION 3. *Does Theorem* 1 *hold when the upper and lower functions are not necessarily constant? What about for Chebyshev systems other than the polynomials? (The techniques in this paper do not seem to work in those cases.)*

References

[BR] J. Briggs and L. A. Rubel, On interpolation by non-negative polynomials, *J. Approx. Theory* **30** (1980), 160–168.

[H] A. L. Horwitz, A constructive approach to non-negative polynomial interpolation, *Approx. Theory Appl.* **3** (1987), 25–36.

[T] G. D. Taylor, Approximation by functions having restricted ranges, III, *J. Math. Anal. Appl.* **27** (1969), 241–248.